Influence of recombination on acquisition and reversion of immune escape and compensatory mutations in HIV-1

Epidemics 14 (2016) 11–25

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Influence of recombination on acquisition and reversion of immune escape and compensatory mutations in HIV-1

Pradeep Nagaraja[a,1], Helen K. Alexander[b,1], Sebastian Bonhoeffer[b,*], Narendra M. Dixit[a,*]

[a] Department of Chemical Engineering, Indian Institute of Science, Bangalore, India
[b] Institute for Integrative Biology, ETH Zürich, Universitätsstrasse 16, CH-8092 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form
11 September 2015
Accepted 11 September 2015
Available online 18 October 2015

Keywords:
Multi-locus adaptation
Immune escape
Cytotoxic T lymphocytes
Waiting time
Fitness landscape

ABSTRACT

Following transmission, HIV-1 adapts in the new host by acquiring mutations that allow it to escape from the host immune response at multiple epitopes. It also reverts mutations associated with epitopes targeted in the transmitting host but not in the new host. Moreover, escape mutations are often associated with additional compensatory mutations that partially recover fitness costs. It is unclear whether recombination expedites this process of multi-locus adaptation. To elucidate the role of recombination, we constructed a detailed population dynamics model that integrates viral dynamics, host immune response at multiple epitopes through cytotoxic T lymphocytes, and viral evolution driven by mutation, recombination, and selection. Using this model, we compute the expected waiting time until the emergence of the strain that has gained escape and compensatory mutations against the new host's immune response, and reverted these mutations at epitopes no longer targeted. We find that depending on the underlying fitness landscape, shaped by both costs and benefits of mutations, adaptation proceeds via distinct dominant pathways with different effects of recombination, in particular distinguishing escape and reversion. When adaptation at a single epitope is involved, recombination can substantially accelerate immune escape but minimally affects reversion. When multiple epitopes are involved, recombination can accelerate or inhibit adaptation depending on the fitness landscape. Specifically, recombination tends to delay adaptation when a purely uphill fitness landscape is accessible at each epitope, and accelerate it when a fitness valley is associated with each epitope. Our study points to the importance of recombination in shaping the adaptation of HIV-1 following its transmission to new hosts, a process central to T cell-based vaccine strategies.

1. Introduction

Cytotoxic CD8+ T lymphocytes (CTLs) mount a powerful response to transmitted HIV in the acute phase of infection and limit its spread within infected individuals (Walker and McMichael, 2012). The CTL response, however, is inadequate to eradicate the virus. HIV evolves to escape this CTL pressure and invariably establishes persistent infection. Understanding the forces driving this arms race between HIV and the early immune response is crucial to the design of preventive vaccines (McMichael et al., 2010; Walker and McMichael, 2012). Recent advances in genomic sequencing technologies have yielded unprecedented insights into the nature of this arms race (Keele et al., 2008; Goonetilleke et al., 2009; Salazar-Gonzalez et al., 2009; Fischer et al., 2010; Henn et al., 2012; Liu et al., 2013). First, in a majority of infected individuals, a single founder strain of the virus is observed to establish infection (Keele et al., 2008; Salazar-Gonzalez et al., 2009; Fischer et al., 2010; Henn et al., 2012). Second, multiple epitopes on the virus are targeted by CTLs, each epitope typically by a distinct CTL clone (Goonetilleke et al., 2009; Henn et al., 2012; Liu et al., 2013). Third, the diversity of the virus in regions targeted by CTLs increases significantly compared to regions not targeted by CTLs, presenting evidence that early viral evolution is in response to CTL pressure (Henn et al., 2012; Liu et al., 2013). Fourth, immune escape mutations accumulate and get fixed at many of these epitopes, subverting CTL activity (Goonetilleke et al., 2009; Henn et al., 2012). Fifth, some regions not targeted by CTLs also evolve (Goonetilleke et al., 2009; Henn et al., 2012; Liu et al., 2013). Since different individuals, depending on

* Corresponding authors.
E-mail addresses: sebastian.bonhoeffer@env.ethz.ch (S. Bonhoeffer), narendra@chemeng.iisc.ernet.in (N.M. Dixit).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.epidem.2015.09.001

their HLA types, target different HIV-1 epitopes, this latter evolution often represents reversion of immune escape mutations in the transmitted strain at epitopes targeted in the donor but not in the recipient (Leslie et al., 2004; Crawford et al., 2007; Chopera et al., 2008; Schneidewind et al., 2009). Immune escape typically incurs a fitness cost, which such reversions serve to recover (Friedrich et al., 2004; Leslie et al., 2004). The emerging paradigm of early HIV adaptation is thus one of rapid, multi-site, targeted evolution on a landscape determined by a convolution of intrinsic viral replicative fitness and CTL pressure.

The large mutation rate of HIV coupled with its massive turnover rate allows the rapid emergence of mutant genomes that could drive early viral adaptation (Coffin, 1995; Perelson et al., 1996; Nowak and May, 2000). Two factors, however, pose limits on such mutation-driven evolution. First, at each epitope, evolution may have to pass through a fitness valley. The mutation on the epitope that confers immune escape often compromises the intrinsic fitness of the virus (Fernandez et al., 2005; Martinez-Picado et al., 2006; Schneidewind et al., 2007; Goepfert et al., 2008; Troyer et al., 2009; Crawford et al., 2009; Miura et al., 2009a,b, 2010). A second, compensatory mutation may arise to restore this fitness at least in part (Kelleher et al., 2001; Brockman et al., 2007; Crawford et al., 2007, 2011; Schneidewind et al., 2007, 2008; Wright et al., 2012; Liu et al., 2014). Such compensatory mutations are however also expected to lower viral fitness when alone (Schneidewind et al., 2007; Althaus and De Boer, 2008; Hinkley et al., 2011; Wright et al., 2012), though in some cases they may be neutral (Liu et al., 2014). That is, the escape and compensatory mutations exhibit positive epistasis and confer a fitness advantage together, but strains carrying either mutation alone may be poorly competitive against the non-mutant strain. Consequently, the frequencies of these single mutant strains are low, limiting the occurrence of the second mutation on these strains. Second, because multiple epitopes are simultaneously targeted by CTLs, the accumulation of the escape and compensatory mutations at each of these epitopes onto a single viral genome is inhibited by 'clonal interference' in finite populations (Gerrish and Lenski, 1998; da Silva, 2012; Ganusov et al., 2013; Kessinger et al., 2013; Leviyang, 2013; Garcia and Regoes, 2014). Clonal interference is illustrated by the hypothetical scenario where CTLs target two epitopes, and viral genomes carrying escape and compensatory mutations to each but not both epitopes exist. Complete escape from CTL pressure would require accumulation of these mutations on a single genome. Until then, however, competition between the strains tends to drive the less fit strain extinct. A study of acutely infected HIV patients revealed a median of eight epitopes targeted by CTLs 2–3 months after the onset of symptoms (Turnbull et al., 2009). Under the simple assumption of one escape and one compensatory mutation per epitope, this would suggest that ~16 mutations must accumulate for optimal adaptation of HIV to the CTL response within the first few months, and, if the previous host targeted completely different epitopes, potentially an equivalent number of reversions. Given the intrinsic mutation rate of HIV in vivo and the size of the infected cell pool, genomes carrying more than two mutations on the fittest genome are unlikely to exist in the viral quasispecies (Gao and Wang, 2007; Gadhamsetty and Dixit, 2010; Pennings, 2012), requiring such accumulation to occur sequentially, with intermediate strains left vulnerable to immune attack. The waiting time for the sequential accumulation of all of these mutations may thus be large. Recombination, which can bring together mutations on distinct viral genomes, may allow HIV to overcome these limitations and adapt rapidly. That recombination can help overcome clonal interference has been recognized earlier as the Fisher–Muller effect (e.g., see Kim and Orr, 2005).

Recombination occurs during reverse transcription of viral RNA in an infected cell when the HIV enzyme reverse transcriptase switches templates between the two strands of the viral RNA contained in the infecting virion and produces a proviral DNA that is a mosaic of the two strands. The crossover frequency of the HIV reverse transcriptase is ~10 per genome per reverse transcription, an order of magnitude larger than the HIV mutation rate (Levy et al., 2004; Suryavanshi and Dixit, 2007). Specifically, if one of the two strands contains an escape mutation on one epitope and the other on another epitope, recombination can yield a genome carrying escape mutations on both epitopes. Similarly, recombination has been shown to accelerate the accumulation of drug resistance mutations (Moutouh et al., 1996). However, the benefit of recombination to HIV is not clear for the following reasons. First, just as recombination can bring mutations together, it can also drive them apart. In particular, when fitness interactions between mutations exhibit positive epistasis, recombination is expected to decelerate the growth of strains carrying multiple mutations (Bonhoeffer et al., 2004; Bretscher et al., 2004). Second, recombination can influence genomic diversification only when the two genomic strands in a virion are not identical; i.e., the virion is heterozygous (Rhodes et al., 2003). The production of heterozygous virions requires cellular superinfection, which appears not to be frequent: recent studies suggest that ~10% of infected cells carry more than one provirus (Josefsson et al., 2011, 2013), arguing for a limited role of recombination (Neher and Leitner, 2010; Batorsky et al., 2011). The influence of recombination then depends on several factors including the frequency of multiple infections of cells, stochastic effects from finite populations, epistatic interactions between escape mutations, and the number of loci involved (Boerlijst et al., 1996; Otto and Lenormand, 2002; Bretscher et al., 2004; Althaus and Bonhoeffer, 2005; Bocharov et al., 2005; Fraser, 2005; Rouzine and Coffin, 2005, 2010; Kouyos et al., 2006, 2007; Gheorghiu-Svirschevski et al., 2007; Hartl and Clark, 2007; Dixit, 2008; Vijay et al., 2008; Arora and Dixit, 2009; Batorsky et al., 2011; Mostowy et al., 2011; Moradigaravand et al., 2014). Broadly, recombination can differentially influence the dynamics of adaptation and the equilibrium frequencies of mutant strains (reviewed, for example, in Dixit, 2008). It is a basic result of population genetics that recombination lowers the magnitude of linkage disequilibrium. In a two-locus/two-allele model, when recombination lowers the extent of negative linkage disequilibrium, the equilibrium frequency of double mutants rises and increases genomic diversity. The opposite happens when recombination lowers positive linkage disequilibrium. Linkage disequilibrium depends on epistatic interactions between loci and on random genetic drift determined by the effective population size. Specifically, random genetic drift introduces negative linkage disequilibrium (the Hill–Robertson effect (Hill and Robertson, 1966)), allowing recombination to increase diversity with small effective population sizes. With large population sizes, linkage disequilibrium tends to have the same sign as epistasis. Moreover, in the two-locus/two-allele model, recombination invariably expedites the first appearance of the strain that carries both mutations, but may decelerate its subsequent growth as described above. With more than two loci, the influence of recombination is thus more difficult to predict: recombination may either accelerate or decelerate the emergence and fixation of higher-order mutants depending on its influence on the emergence and growth of lower-order mutants. The conditions under which recombination accelerates adaptation at multiple loci thus remain to be clarified.

Several previous studies have used mathematical models to investigate whether recombination is expected to accelerate or decelerate the emergence of multiple drug resistance in HIV, with inconsistent conclusions (Bretscher et al., 2004; Althaus and Bonhoeffer, 2005; Fraser, 2005; Carvajal-Rodriguez et al., 2007; Kouyos et al., 2009). The most recent of these suggested that both population dynamics (changing population size) and stochasticity (via the Fisher–Muller effect) have important influences on

the results, clarifying the previous inconsistencies and suggesting that recombination tends to have little if any accelerating effect on the emergence of resistance, especially when positive epistasis between mutations is taken into account (Kouyos et al., 2009). The aforementioned studies considered two-locus/two-allele models. However, a recent study considering a large number of loci with fitness parameterized empirically suggested that recombination could accelerate adaptation of HIV over broad parameter ranges (Moradigaravand et al., 2014). A recent modelling study specifically considering the role of recombination in immune escape found that recombination tends to accelerate escape, with the effect increasing with the number of loci involved (Mostowy et al., 2011). The latter study suggested that recombination could have a greater role when the cost of escape is smaller, but the main results illustrated an average over randomly-drawn fitness costs (Mostowy et al., 2011). In the present study, we extend investigations of the role of recombination in the adaptation of HIV to a host's CTL response, by considering in detail the influence of the fitness landscape. Moreover, we consider both escape at targeted epitopes and reversion at non-targeted epitopes, which present distinct features of the fitness landscape. We investigate these questions using a population dynamics model that treats co-infection and recombination mechanistically, and accounts for finite population effects through a waiting time for new genotypes to appear.

Our model builds on previous studies that explicitly consider the processes of cellular superinfection, tempered by CD4 down-modulation; the production of novel viral genotypes via mutation and recombination; and/or the dynamics and evolution of the virus under immune-driven selection (Bretscher et al., 2004; Althaus and Bonhoeffer, 2005; Suryavanshi and Dixit, 2007; Vijay et al., 2008; Arora and Dixit, 2009; Gadhamsetty and Dixit, 2010; Mostowy et al., 2011; van Deutekom et al., 2013; Batorsky et al., 2014). We use a modelling formalism similar to previous studies that were able to quantitatively capture in vitro dynamics of the growth of recombinant strains (Suryavanshi and Dixit, 2007) and the development of resistance to drugs with large genetic barriers (Arora and Dixit, 2009). Here, we advance the formalism to include the influence of CTLs in vivo and apply it to examine the role of recombination in immune escape and reversion. We compute the expected waiting times for the emergence of strains carrying highly adapted combinations of escape and reversion mutations. We find that recombination may accelerate or decelerate this emergence, depending on the number of epitopes involved in the adaptation process and on the fitness landscape, as determined by immune pressure and the cost of escape.

2. Model

We construct a mathematical model to describe within-host HIV-1 evolution in the context of immune pressure from CTLs. The model consists of a system of ordinary differential equations describing the dynamics of all viral strains that are present, combined with an expected waiting time until mutant strains appear, derived from a stochastic process approach. This modelling approach is based on the multi-strain viral dynamics model presented in Arora and Dixit (2009), extended to take into account the host's immune response.

We consider n epitopes on the viral genome undergoing adaptation (Fig. 1). 'Targeted' epitopes are those newly recognized by the immune response in the new host, while 'reverting' epitopes are no longer recognized and thus undergoing reversion of mutations fixed in the previous host. On each epitope, we consider two positions at which mutations affect viral fitness, one an escape mutation and the other a compensatory mutation. Thus, a total of 2^{2n} different viral variants can arise of which $S = 2^{2n} - 1$ are mutants

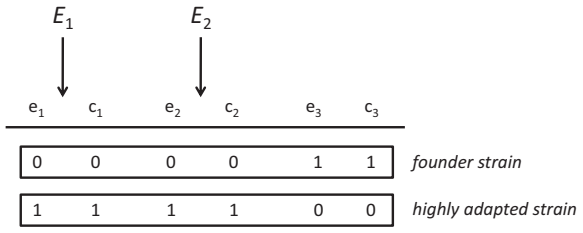

Fig. 1. Schematic of early HIV-1 adaptation. Illustration showing a viral sequence consisting of three epitopes with escape mutations e_1, e_2 and e_3 and corresponding compensatory mutations c_1, c_2 and c_3. Epitopes 1 and 2 are targeted by CTL clones E_1 and E_2, whereas epitope 3 is not targeted in this host. The corresponding founder sequence and the highly adapted strain are shown with mutations indicated by '1' and wild type alleles indicated by '0'.

and one is the wild-type. We denote the different strains with the index i where $i \in \{0, 1, \ldots, S\}$. We let infection begin with the founder strain that carries no mutations on the targeted epitopes and both mutations on each of the reverting epitopes. We assume that the most highly adapted strain is the one that carries both mutations on all the targeted epitopes and none on the reverting epitopes. (Note that both the founder and highly adapted strains are generally distinct from the 'wild type' which carries no mutations.) We construct dynamical equations to estimate the time of emergence of this highly adapted strain.

The equations track the temporal dynamics of populations of uninfected target cells, T; cells infected with a single provirus i, T_i; cells infected with two proviruses i and j, T_{ij}; free virions carrying genomes i and j, V_{ij}; and CTL effector cells of clone a, E_a, where $i \in \{0, 1, \ldots, S\}$, $j \in \{i, i+1, \ldots, S\}$ and $a \in \{1, 2, \ldots, m\}$:

$$\frac{dT}{dt} = \lambda - d_T T - k_0 T \sum_{i=0}^{S} H(t - w_i) q_i$$

$$\frac{dT_i}{dt} = H(t - w_i) k_0 T q_i - k_1 T_i \left(\sum_{j=0}^{i-1} H(t - w_{ji}) q_j + \sum_{j=i}^{S} H(t - w_{ij}) q_j \right) - \delta T_i - k_{CTL} T_i \sum_{a=1}^{m} \frac{\alpha_{ia} E_a}{h + G_{ii} + U_a}$$

$$\frac{dT_{ii}}{dt} = H(t - w_{ii}) k_1 T_i q_i - \delta T_{ii} - k_{CTL} T_{ii} \sum_{a=1}^{m} \frac{2\alpha_{ia} E_a}{h + G_{ii} + U_a}$$

$$\frac{dT_{ij}}{dt} = H(t - w_{ij}) k_1 (T_i q_j + T_j q_i) - \delta T_{ij} - k_{CTL} T_{ij} \sum_{a=1}^{m} \frac{(\alpha_{ia} + \alpha_{ja}) E_a}{h + G_{ij} + U_a} \qquad (i \neq j)$$

$$\frac{dV_{ii}}{dt} = p \left(T_i + T_{ii} + \frac{1}{4} \sum_{j=0}^{i-1} T_{ji} + \frac{1}{4} \sum_{j=i+1}^{S} T_{ij} \right) - c V_{ii}$$

$$\frac{dV_{ij}}{dt} = \frac{p}{2} T_{ij} - c V_{ij} \qquad (i \neq j)$$

$$\frac{dE_a}{dt} = \sigma - \delta_E E_a + g E_a \sum_{i=0}^{S} \frac{\alpha_{ia} T_i}{K_a + G_{ii} + U_a} + g E_a \sum_{j=0}^{S} \sum_{k=j}^{S} \frac{(\alpha_{ja} + \alpha_{ka}) T_{jk}}{K_a + G_{jk} + U_a}$$

where we denote the total rate of infection producing provirus i as

$$q_i := \sum_{j=0}^{S} \sum_{k=j}^{S} Q_i(jk) V_{jk} f_{jk} ;$$

the total population of CTLs recognizing strain i and/or j as

$$G_{ij} := \sum_{a=1}^{m} (\alpha_{ia} + \alpha_{ja} - \alpha_{ia}\alpha_{ja}) E_a ;$$

and the total population of infected cells (singly or doubly) containing at least one provirus recognized by effector clone a as

$$U_a := \sum_{j=0}^{S} \alpha_{ja} T_j + \sum_{j=0}^{S} \sum_{k=j}^{S} (\alpha_{ja} + \alpha_{ka} - \alpha_{ja}\alpha_{ka}) T_{jk} .$$

Here, uninfected CD4+ target cells, T, are produced at the rate λ and die at the rate d_T. Further, uninfected cells are lost to infections by virions V_{jk} at the infection rate constant k_0 scaled by the relative infectivity f_{jk}. After infection of a target cell, a virus of type jk mutates and recombines during the process of reverse transcription producing provirus i with a probability $Q_i(jk)$ (see below).

Singly infected cells are produced from infections of uninfected cells and are lost to double infections with the rate constant k_1, again scaled by relative infectivity. They die at the rate δ per day and are lost due to CTL killing, by recognition of epitopes on their surfaces, according to a saturating rate (see below). CTLs recognizing epitope a, E_a, kill infected cells T_i, T_{ii} and T_{ij} that expose such epitopes on their surface. The entries of the binary recognition matrix, α_{ia}, are 1 if E_a recognizes T_i, otherwise 0. We assume that the killing efficiency is the same for any CTL clone once its target epitope is recognized.

In a finite population, the formation of the infected cell T_i is stochastic as the rate of infection leading to these cells can be small for a genome i several mutational steps away from the founder strain. We determine the time to emergence, t_i, of the first cell T_i assuming that infections occur according to a time-inhomogeneous Poisson process. We denote the expectation of this time to emergence as w_i. We assume, as an approximation, that T_i emerges at its expected waiting time, w_i (Arora and Dixit, 2009). Thus, if the time elapsed from the start of this adaptation process, t, is less than w_i, then T_i is yet to be formed, which is captured by the Heavyside function: $H(t - w_i) = 0$ if $t < w_i$, otherwise $H(t - w_i) = 1$. A similar condition applies to the doubly infected cells. The evaluation of waiting times is described below.

Doubly infected cells are produced from singly infected cells with the rate constant k_1, lower than k_0, as subsequent to single infection, CD4 receptors are down-modulated reducing the susceptibility to new infection (Chen et al., 1996; Piguet et al., 1999). If the same provirus i infects the singly infected cell T_i, then the doubly infected cell T_{ii} is homozygous. If a different provirus j infects the singly infected cell T_i (or i infects T_j), then the doubly infected cell T_{ij} is heterozygous. We ignore more than two infections per cell following previous studies (Suryavanshi and Dixit, 2007; Arora and Dixit, 2009; Batorsky et al., 2011; Josefsson et al., 2013). We assume next that doubly infected cells T_{ii} present epitopes to twice the extent of singly infected cells T_i, rendering T_{ii} twice as susceptible to CTL killing as T_i. Similarly, the susceptibility of cells T_{ij} to immune killing equals the sum of that of T_i and T_j. (Modifying this assumption does not alter our qualitative conclusions; see Fig. S1.)

While cells of type T_i and T_{ii} produce only homozygous virions V_{ii}, cells T_{ij} produce virions V_{ii}, V_{jj}, and the heterozygous virions V_{ij} in the proportion of 1/4, 1/4, 1/2, representing random assortment from a pool of strands i and j (Chen et al., 2009). Infected cells produce virions at rate p, and virions are cleared at rate c. (We assume that singly and doubly infected cells produce virions at the same rate, but again, qualitative conclusions are robust to modifying this assumption; see Fig. S1.)

Effector cells E_a are produced at an intrinsic rate σ and proliferate following their interactions with infected cells at a saturating rate (see below). The distribution of the saturation parameter K_a would define the immunodominance pattern of the effector response (Mostowy et al., 2011); however, for our results we will take K_a the same for all clones a. The effector cells die at the rate δ_E.

In both CTL proliferation and killing, we employ saturating functional forms, which reflect the limitations in the rate of cellular interactions, for instance due to the time to find interacting partners and the handling time spent in each interaction (Miller et al., 2004; Beauchemin et al., 2007; Graw and Regoes, 2009; Gadhamsetty et al., 2014). These considerations imply that interactions cannot occur at an arbitrarily high rate even as both cellular populations (infected cells and CTLs) become large, unlike in a mass action functional form. That saturating functional forms are more biologically realistic has been argued in a recent review on modeling T cell dynamics (Wodarz, 2014). However, mass action, for instance as observed in recent studies (Budhu et al., 2010; Ganusov et al., 2011), can be recovered in the limit as both maximal rates and saturation parameters become large, i.e., both k_{CTL} and h in the CTL killing term or both g and K_a in the CTL proliferation term. The functional forms we use can be derived from kinetic considerations of complex formation and dissociation and making use of the "total quasi-steady state assumption" (Borghans et al., 1996; De Boer, 2007). Here, saturation occurs with respect to the populations of both interacting partners. We note that previous models of immune escape have also made use of saturating functional forms in CTL proliferation (De Boer, 2007; Althaus and De Boer, 2008; Mostowy et al., 2011; van Deutekom et al., 2013; Batorsky et al., 2014) and/or killing (De Boer, 2007; Althaus and De Boer, 2008; Gadhamsetty et al., 2014).

We estimate the values of the parameters involved based on available experimental/clinical observations and previous modelling studies (see Text S1). The parameter values we use are listed in Table 1. We solve the above equations numerically with the fourth order Runge–Kutta algorithm using a computer program in C.

2.1. Reverse transcription

We now describe how we calculate the probability of producing a given provirus during reverse transcription of a given double-stranded infecting virion. For this purpose, we follow previous authors in decoupling mutation and recombination (Boerlijst et al., 1996; Bocharov et al., 2005; Arora and Dixit, 2009). The procedure closely follows that developed in (Arora and Dixit, 2009). The probability of producing genome k from genomes j and h via recombination is denoted as $R_k(jh)$, while the probability of producing genome i via mutation from the recombinant k is denoted by P_{ik}. In all, 2^d different recombinants k, where d is the number of sites at which j and h differ, can be produced. Summing over the product of the probability of producing any recombinant k and the probability of mutating k to i yields the total probability of producing genome i by reverse transcription of genomes j and h,

$$Z_i(jh) = \sum_{k=0}^{2^d-1} P_{ik} R_k(jh)$$

Whereas the above is the probability of producing a strain i following reverse transcription, we recognize in our dynamical equations that such a strain may only arise after its waiting time is crossed. Thus, we determine a normalized probability by restricting the appearance of proviruses to those genomes whose waiting times have been crossed, obtaining

$$Q_i(jh) = \frac{H(t - w_i) Z_i(jh)}{\sum_{k=0}^{S} H(t - w_k) Z_k(jh)}$$

2.1.1. Recombination

We begin with genomes j and h that differ at d sites separated by distances (number of base pairs) l_1, l_2, …, l_{d-1}. The enzyme

Table 1
Parameter values used for simulation.

Parameter	Description	Value
λ	Influx rate of target cells	2×10^7 cells/day
d_T	Per capita death rate of uninfected target cells	0.05/day
k_0	Infection rate of uninfected cells, per virion and per uninfected cell	10^{-11}/day
k_1	Superinfection rate (of already singly-infected cells)	$0.7k_0$ unless otherwise noted
f_{ij}	Relative infectivity of virus carrying genomes i and j (multiplies k_0 and k_1)	1 for wild type, variable (<1) otherwise
D^*	Total death rate of singly infected cells at steady state, assuming susceptibility to immune killing	1/day
ϵ	Steady state proportion of singly infected cell death due to CTL killing, assuming no escape	Variable
δ	Death rate of infected cells due to sources other than CTL killing at the focal epitope(s)	$D^*(1-\epsilon)$
k_{CTL}	Maximum rate of CTL killing of singly infected cells	Calculated from other parameters (see Suppl. Text S1)
h	Saturation parameter of CTL killing	5×10^8 cells
α	Binary immune recognition matrix	$\alpha_{ia} = 1$ if viral strain i is recognized by CTL clone a, 0 otherwise
σ	Influx rate of CTLs	20 cells/day
g	Proliferation rate constant of CTLs	1/day unless otherwise noted
K_a	Saturation parameter of CTL proliferation for clone a	10^8 cells
δ_E	Per capita death rate of effector cells	0.02/day
p	Per capita production rate of virions by infected cells	5×10^4 virions/day
c	Clearance rate of virions	23/day
μ	Mutation rate	3×10^{-5}/site/replication unless otherwise noted
ρ	Recombination rate	8.3×10^{-4} crossovers/site/replication
n	Number of epitopes	variable
S	Number of non-wild type viral strains (types of genomes)	$S = 2^{2n} - 1$

reverse transcriptase (RT) initiates reverse transcription on either of the two strands with equal probability. Subsequently, RT can cross over genomes and thus combine genetic material from both strands in the process. To generate genome k from genomes j and h, RT has to visit these d sites on the appropriate genomes. If RT performs an even number of crossovers between successive sites, then it remains on the same genome at these sites, otherwise it shifts to the other genome. If $P_{des}(x)$ is the probability that RT is on the desired genome at the x^{th} distinctive site via odd or even crossovers over the length l_{x-1}, then the probability of generating genome k via recombination is

$$R_k(jh) = \prod_{x=1}^{d} P_{des}(x)$$

Based on the desired path to generate genome k, at the first site $P_{des}(1) = 1/2$, as RT can start on either genome with equal probability. If the next desired site on genome k is on the same genome as the residing genome, then RT has to undergo an even number of crossovers and hence $P_{des}(2) = P_{even}(l_1)$. Otherwise, the desired site is on the other genome and RT has to undergo an odd number of crossovers, so that $P_{des}(2) = P_{odd}(l_1)$. The probabilities $P_{even}(l)$ and $P_{odd}(l)$ are given as $(1 + (1-2\rho)^l)/2$ and $(1 - (1-2\rho)^l)/2$ respectively, where ρ is the per site recombination rate of HIV, assuming a binomial distribution of crossover events over a given number of sites (see Text S2).

2.1.2. Mutation

From the genome k derived from recombination of genomes j and h, we obtain the desired genome i by mutation. If genomes i and k differ at u sites, then the probability that we arrive at genome i from genome k is $P_{ik} = \mu^u(1-\mu)^{2n-u}$, where $0 \le u \le 2n$ and μ is the per site mutation probability.

2.2. Waiting times

The time to emergence, t_i, of the first singly infected cell T_i can be modelled by a time-inhomogeneous Poisson process with the instantaneous rate $r_i(t)$ of the formation of provirus i (Arora and Dixit, 2009). The probability that t_i is smaller than s is then

$$P(t_i \le s) = 1 - \exp\left(-\int_0^s r_i(t)dt\right), \quad 0 \le s < \infty$$

The expected waiting time for the formation of the first singly infected cell of type i, $w_i = \langle t_i \rangle = \int_0^\infty s(\partial P/\partial s)ds$ or

$$w_i = -\int_0^\infty s\frac{\partial}{\partial s}\left(\exp\left(-\int_0^s r_i(t)dt\right)\right)ds$$

Similarly for doubly infected cells, T_{ij}, the expected waiting time for the first formation is

$$w_{ij} = -\int_0^\infty s\frac{\partial}{\partial s}\left(\exp\left(-\int_0^s r_{ij}(t)dt\right)\right)ds$$

The required rates are given by:

$$r_i(t) = k_0 T(t) \sum_{j=0}^{S}\sum_{k=j}^{S} Z_i(jk)V_{jk}(t)f_{jk},$$

$$r_{ii}(t) = k_1 T_i(t) \sum_{j=0}^{S}\sum_{k=j}^{S} Z_i(jk)V_{jk}(t)f_{jk},$$

$$r_{ij}(t) = k_1\left(T_i(t)\sum_{k=0}^{S}\sum_{l=k}^{S} Z_j(kl)V_{kl}(t)f_{kl} + T_j(t)\sum_{k=0}^{S}\sum_{l=k}^{S} Z_i(kl)V_{kl}(t)f_{kl}\right) \quad \text{for } i < j.$$

To evaluate the above integrals for w_i and w_{ij}, we follow the numerical procedure developed in (Arora and Dixit, 2009). This procedure dynamically updates the estimated w_i and w_{ij} at each time step in the numerical integration of the ODEs, with the rates calculated from current population sizes in the system.

3. Results

We initialize the system to the steady state where only the founder strain is present (see Text S1), and focus on the effect of recombination on the waiting time until the highly adapted strain

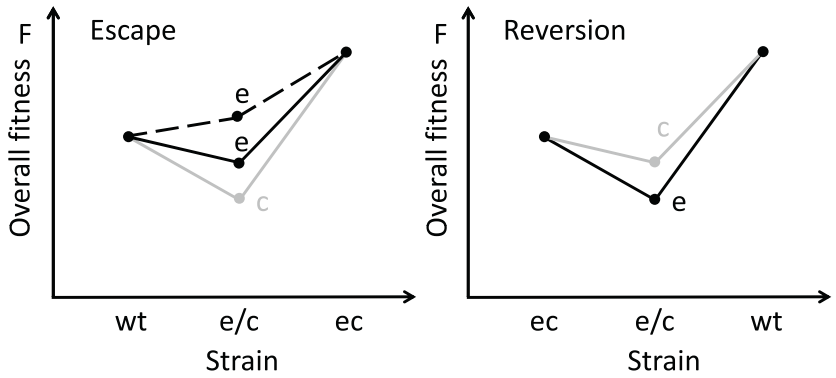

Fig. 2. Fitness landscapes at one epitope. The overall fitness (F) of strains carrying no mutations (wt), an escape mutation (e), a compensatory mutation (c), and both escape and compensatory mutations (ec) are shown. Adaptation can pass through a fitness valley (solid) or a purely uphill trajectory (dashed) during escape (left), but must pass through a valley during reversion (right).

first appears. This effect is quantified by $R = (w_{rec} - w_{norec})/w_{norec}$, which is the relative change in this waiting time "with" versus "without" recombination. Thus $R < 0$ indicates that recombination has an accelerating effect, while $R > 0$ indicates a decelerating effect on adaptation. From an evolutionary theoretical perspective, a comparison between non-zero crossover probability and the hypothetical scenario of zero crossover probability allows evaluation of the benefit of recombination to the virus. From a genetics perspective, the comparison allows investigation of the effect of genome architecture – characterized by the genetic distance between loci – on viral adaptation, given that it undergoes recombination. In practice, for numerical computations, we leave the crossover probability (ρ) fixed and modify the inter-locus distances. In the results that follow, we compare approximately free recombination (i.e. a large distance) between all or selected loci, to the baseline of complete linkage (zero distance) between all loci. Note that the ordering of loci does not matter in the two extreme cases of complete or no linkage between all loci.

3.1. Single epitope

We first consider immune escape and reversion at a single epitope. The pathway of adaptation depends strongly on the underlying fitness landscape, which combines intrinsic fitness (mediated in our model by the relative infectivity, f) and susceptibility to immune attack to yield the overall or effective fitness, denoted F. The effective fitness landscape of the virus, determining the direction and strength of selection, depends on the current availability of target cells and the presence of CTLs recognizing this epitope, and will thus generally show complex temporal dynamics (see Text S1). For intrinsic fitness, we assume $f_{wt} > f_{ec} > f_e, f_c$; that is, the wild type strain has the highest relative infectivity, while e and c mutations come with costs but demonstrate positive epistasis (compensation). CTLs recognizing a strain affect its survival. If D^* is the steady state total death rate of singly infected cells and δ the background death rate from sources other than killing by CTLs targeting the focal epitope, then $\epsilon = 1 - \delta/D^*$ is the proportion of infected cell death due to killing by CTLs targeting this epitope. Since the total death rate of infected cells is better known than the relative contributions of background death rate and immune killing (see Text S1), we fix D^* for an immune-susceptible strain, and vary the proportion of death due to immune killing, ϵ. The parameter ϵ thus mediates the fitness advantage that can be gained by immune escape.

In the escape scenario, where CTLs targeting the epitope are present, the founder strain is the wild type carrying no mutations, denoted wt, and the highly adapted strain is ec, carrying both the escape, e, and compensatory, c, mutations. That is, $F_{ec} > F_{wt}$. Further, $F_c < F_{wt}$, representing the fitness penalty associated with the compensatory mutation alone. Although $f_e < f_{wt}$, indicating the intrinsic fitness loss associated with the escape mutation, F_e could be greater or smaller than F_{wt} depending on the strength of immune killing. If the proportion of death due to immune killing, ϵ, is large, then F_e is likely to be larger than F_{wt}, owing to the large fitness gain associated with immune escape. If ϵ is small, however, F_e is likely to be smaller than F_{wt}. Thus, depending on ϵ and on f, and hence on whether F_e is smaller or larger than F_{wt}, the immune escape pathway either must pass through a fitness valley or can access a purely uphill trajectory (Fig. 2A).

In the reversion scenario, CTLs do not target the focal epitope; thus adaptation occurs under fixed immune pressure on all strains, which we suppose is due to CTLs targeting other epitopes (e.g., see Pandit and de Boer, 2014). Thus, the total death rate is the same for all strains, and overall fitness is mediated by relative infectivity. Adaptation therefore always passes through a fitness valley (F_e, $F_c < F_{ec} < F_{wt}$) (Fig. 2B). Note that we still use e and c to refer to the mutations, although e does not confer immune escape in the new host.

Below, we examine the role of recombination in the adaptation process under each of these circumstances. We consider a range of intrinsic fitness landscapes defined by the parameters f_e and f_c. In the one-epitope case, the fully adapted strain is two mutational steps from the founder, and recombination, by potentially bringing together single-mutants, can thus only have an accelerating effect on its first appearance. That is, the direction of the effect is clear, but the question is its magnitude. Note that once the highly adapted strain appears, however, recombination can slow its spread by breaking up the two mutations again, though we do not focus on this phase.

3.1.1. Escape

During escape, from the wt strain present initially, the single mutant genomes e and c are produced by mutations at the same expected waiting times, as their production rates depend only on the replication rate of the wt and the mutation rate, and not on the relative fitness values of the mutants (Figs. 3A and S2). Following its emergence, the strain c is competitively suppressed by wt, because $F_c < F_{wt}$, and thus settles to a low frequency. When f_e and/or ϵ are sufficiently large such that $F_e > F_{wt}$, the strain e has a selective advantage and can outcompete the wt. It thus rises rapidly following its emergence. When, as a consequence, the e strain becomes dominant, the rate at which the ec strain arises by mutation of the e strain is much larger than by recombination between the e and c strains. Recombination, therefore, plays a negligible role in the emergence of the ec strain under these circumstances. That is, the expected waiting time of the highly adapted (ec) strain with and without recombination is similar, $w_{rec} \approx w_{norec} \approx 5$ days ($f_e = 0.8$ in

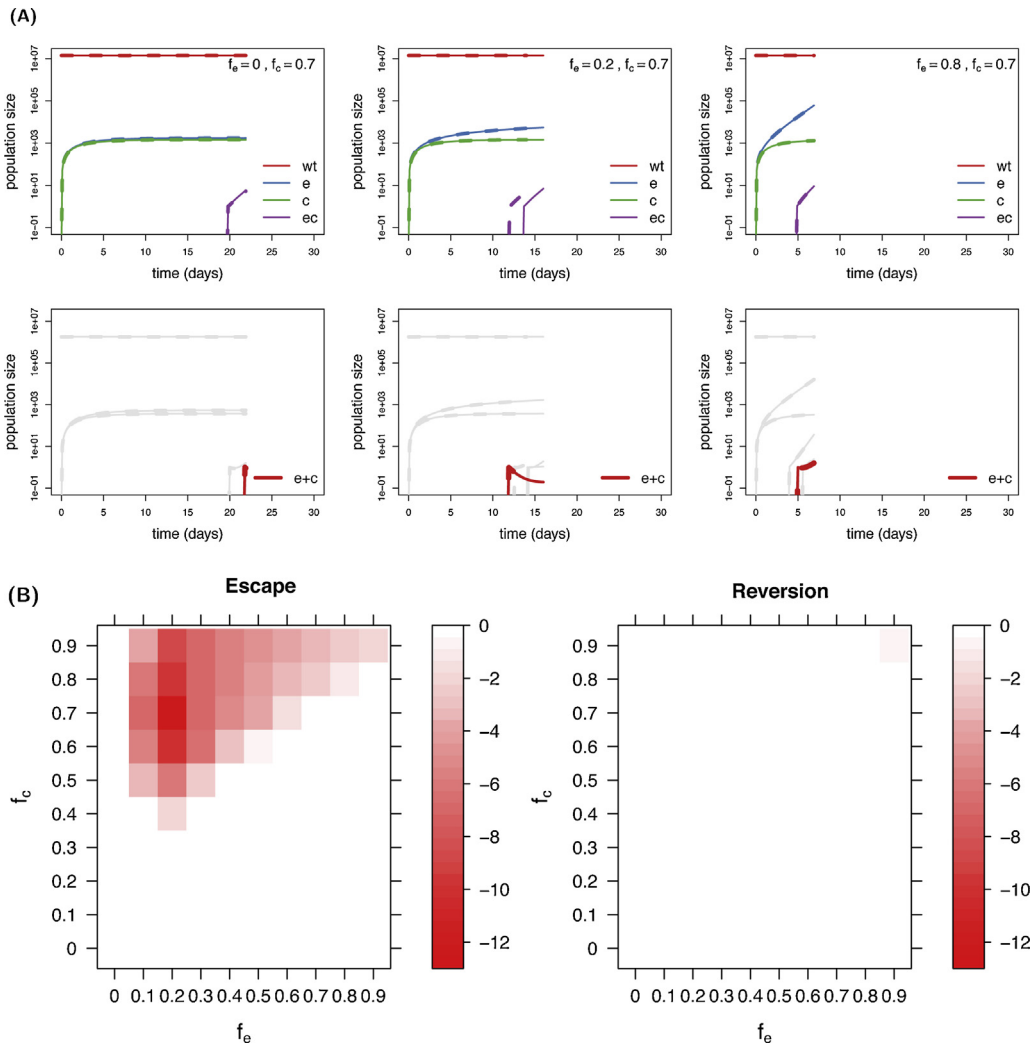

Fig. 3. Escape and reversion at a single epitope. We set the intrinsic fitness of the founder strain (f_{wt}) to 1 and the highly adapted strain (f_{ec}) to 0.9 throughout, and vary the intrinsic fitness associated with escape (f_e) and compensatory (f_c) mutations. The relative contribution of immune killing is $\epsilon = 0.9$. Other parameter values are listed in Table 1. (A) Temporal dynamics of escape at a single epitope for varying intrinsic fitness values of the escape strain (given by f_e in the legend on each panel). Populations of singly infected cells with proviruses *wt*, *e*, *c* and *ec* (top panel) and populations of doubly infected cells carrying proviruses *e* and *c* (red) and various other pairs of proviruses (grey) (bottom panel). Solid lines indicate dynamics without recombination and dashed lines with recombination. (B) The percentage change in the expected waiting time with versus without recombination ($R * 100\%$) for varying intrinsic fitness values of escape and compensatory mutations for the cases of escape (left) and reversion (right). (For interpretation of reference to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3A), so that our measure of the influence of recombination, $R = (w_{rec} - w_{norec})/w_{norec}$, is nearly zero. If, on the other hand, f_e is smaller, so that $F_e \approx F_{wt}$, then *e* does not have a significant advantage over *wt*, but the rate at which it superinfects cells infected by *c* remains significant. Then recombination can accelerate the accumulation of the two mutations and hence the emergence of the *ec* strain ($w_{rec} \approx 12$ days and $w_{norec} \approx 13.5$ days so that $R \approx -12\%$, when $f_e = 0.2$ in Fig. 3A). If f_e is too low, then *e* settles to such a low frequency that coinfection of cells by *e* and *c* becomes rare, and recombination again does not accelerate escape ($w_{rec} \approx w_{norec} \approx$ 20 days, so that $R \approx 0\%$ when $f_e = 0$ in Fig. 3A). Similar trends are observed as f_c is altered for fixed f_e. Fig. 3B illustrates how R varies across the full range of f_e and f_c.

We examine next the influence of parameter variations on the effect of recombination, as quantified by R. Increasing the frequency of doubly infected cells, by increasing the superinfection rate k_1 of singly infected cells, enhances the effect of recombination (Fig. S3). For smaller mutation rates (μ), the relative influence of recombination is also larger (Fig. S4A), while for large values of μ, the influence of recombination is overwhelmed by mutation and R is zero across the fitness landscape (Fig. S4B). Decreasing the relative contribution of immune killing to infected cell death (ϵ) lowers the fitness advantage associated with escape and hence the uphill immune escape pathway only becomes accessible at higher intrinsic fitness of the escape strain. Thus, while the qualitative pattern described above remains similar, the accelerating effect of recombination is evident at correspondingly higher values of f_e when ϵ is smaller (Fig. S5).

3.1.2. Reversion

We next consider the reversion of an escape mutant in a new host. Here, we begin with the *ec* strain. Both the single mutant *e* and *c* strains again arise due to mutation of the *ec* strain. Because the fitness landscape always passes through a valley, they both remain at low frequencies relative to the *ec* strain. The rate at which cellular superinfection with the *e* and *c* strains occurs is thus substantially smaller than superinfection with the *ec* strain. Consequently, the production of heterozygous virions, which provide the substrate for recombination, is minimal. R thus remains nearly zero throughout the fitness landscape, until f_e and f_c become comparable

to f_{ec}, which here is set to 0.9 (Fig. 3B). Recombination thus does not have a significant effect on reversion associated with a single epitope. This result is robust to variations in the parameters k_1, μ, and ϵ (Figs. S3–S5).

3.2. Two epitopes

We now consider adaptation at two epitopes. At each epitope, we let mutations affect intrinsic fitness as defined in the single epitope case above. Thus, f_{e_1} is the intrinsic fitness of the single mutant strain carrying the escape mutation on the first epitope. For simplicity, we assume that the intrinsic fitness landscapes governing the two epitopes are identical, so that $f_{e_1} = f_{e_2}$, and so on. We thus isolate the effect of simply having more sites at which adaptive changes are required. Further, we assume that the intrinsic fitness effects exhibit no inter-epitope epistatic interactions, so that, e.g., $f_{e_1e_2} = f_{e_1}f_{e_2}$.

To examine the influence of recombination, we consider two arrangements of loci. In the first, the escape and compensatory mutations are all sufficiently far apart on the genome to allow maximal recombination between all loci. In the second, the escape and compensatory mutations associated with a given epitope are proximally located but the epitopes are far apart, so that recombination has no influence on adaptation at any one epitope, but can exert its influence across the epitopes. For each scenario, we again explore the influence of recombination across a range of intrinsic fitness landscapes defined by the parameters f_e ($= f_{e_1} = f_{e_2}$) and f_c ($= f_{c_1} = f_{c_2}$). Three cases arise: (1) escape at both epitopes, (2) reversion at both epitopes, and (3) escape at one and reversion at the other epitope.

3.2.1. Escape at both epitopes

In this case, the founder strain is the wild type, *wt*. We examine the effect of recombination on the expected waiting time for the strain that accumulates escape and compensatory mutations at both epitopes, denoted $e_1c_1e_2c_2$, where the subscripts indicate the epitope number.

In contrast to the single epitope results, we find here that recombination can accelerate or decelerate adaptation (measured again by $R = (w_{rec} - w_{norec})/w_{norec}$, where w is the waiting time for the appearance of the highly adapted strain, here $e_1c_1e_2c_2$) depending on the underlying fitness landscape (Fig. 4A). For instance, for the chosen parameter values, as f_e increases with f_c fixed, recombination exhibits a complex, non-monotonic influence. We can understand these effects by again looking at temporal dynamics, illustrated in Fig. 5 for the example where $f_c = 0.4$ and f_e varies from 0.4 to 0.7. Note that by symmetry, the two epitopes are interchangeable, and thus the corresponding strains (e.g. e_1c_1 and e_2c_2) have identical dynamics.

Starting from the *wt* strain, all single mutant strains (e_1, c_1, e_2, and c_2) are produced at the same expected time. The strains however settle to low frequencies given their poor fitness relative to *wt*. Because $F_e > F_c$ in this example (due to the advantage of escaping at one epitope and given that $f_e \geq f_c$ here), the e_1 and e_2 strains are slightly higher in prevalence than c_1 and c_2, and their prevalence increases with f_e across panels. The double mutant e_1e_2 appears next, closely followed by e_1c_1 and e_2c_2. Due to the lower fitness of c, the double mutant c_1c_2 appears somewhat later and remains at very low frequencies, thus playing a negligible role in the dynamics. Though recombination among the corresponding single mutants could potentially play a role in the emergence of the double mutants, it appears that this effect is minimal.

Under the chosen parameter settings ($f_{ec} = 0.9$, combined with the advantage of escape at one epitope), e_1c_1 and e_2c_2 are fitter than *wt* and thus grow. Here, recombination inhibits their growth because recombination predominantly occurs with the wild type, breaking each *ec* into its constituent single mutants. The effect is evident in the significantly delayed emergence of all triple mutant strains in the presence of recombination. Nonetheless, the possibility of recombining e_1c_1 and e_2c_2 to produce the desired quadruple mutant $e_1c_1e_2c_2$ points to a counteracting advantage of recombination.

The net effect of recombination varies with f_e across the panels. (Note that as f_e increases, the quadruple mutant emerges faster in absolute terms, either with or without recombination; however, we focus on the *relative* effect of recombination, as quantified by R, for each f_e.) This variation can be traced to the dynamics of the strains e_1e_2, $e_1c_1e_2$, and $e_1e_2c_2$, which depend on f_e. When $f_e = 0.4$ (Fig. 5A), the aforementioned double mutant does not rise in frequency, and the triple mutants rise relatively slowly due to only a small advantage over *wt*. Recombination plays a significant role in bringing together e_1c_1 and e_2c_2 to accelerate the appearance of $e_1c_1e_2c_2$ ($R \approx -20\%$). When $f_e = 0.5$ (Fig. 5B), the triple mutants rise faster and the rate of production of $e_1c_1e_2c_2$ by one additional mutation is larger. Consequently, the relative role of recombination declines ($R \approx -5\%$). Once $f_e = 0.6$ (Fig. 5C), the double mutant e_1e_2 also becomes significantly fitter than the *wt* and thus rises in frequency. The predominant pathway of adaptation now begins to shift from being dominated by the accumulation of mutations on the e_1c_1 and e_2c_2 backgrounds to that on the e_1e_2 background. Once the triple mutants $e_1c_1e_2$ and $e_1e_2c_2$ appear, recombination can bring them together to produce the quadruple mutant. However, the inhibitory effect of recombination in breaking up the double mutant strains (through recombination with *wt*) is also evident, delaying the emergence of the triple mutants. The net effect of recombination on the emergence of the quadruple mutant becomes inhibitory ($R \approx 10\%$). As f_e increases further to 0.7 (Fig. 5D), the intermediate strains e_1e_2, $e_1c_1e_2$, and $e_1e_2c_2$ all rise rapidly after their appearance, and the emergence of $e_1c_1e_2c_2$ through sequential accumulation of mutations swamps the effect of recombination. The net effect of recombination remains inhibitory here, but is subdued to $R \approx 5\%$.

When free recombination occurs only between the two epitopes while the escape and compensatory mutations associated with each epitope are tightly linked, recombination tends to have a more beneficial (or less inhibitory) effect on the emergence of the quadruple mutant throughout (Fig. 4A). Essentially, this is because we eliminate the detrimental effect of recombination in breaking up e from c at each epitope by recombination with the wild type, while leaving its main benefit of bringing together e_1c_1 and e_2c_2 intact. Only at the highest f_e does sequential mutation swamp recombination and $R \approx 0$.

Decreasing the relative strength of immune killing (ϵ) shifts the balance increasingly toward the inhibitory effects of recombination when recombination is allowed between all loci (Figs. S6A and S6B). As ϵ decreases, and correspondingly the background death rate is higher, the overall fitness values F_e and F_{ec} of strains carrying escape mutations decrease for the same intrinsic fitness values f_e and f_{ec}. This reduced fitness implies that e_1c_1 and e_2c_2 are present at reduced frequencies, and when recombination is allowed, it occurs primarily with the wild type. When ϵ is small, recombination thus inhibits the growth of e_1c_1 and e_2c_2 by breaking up the two combinations (Fig. S7), while co-infection with both e_1c_1 and e_2c_2 is rare and therefore recombination can play only a small role in bringing these two pairs of mutations together. Consequently, recombination decelerates the emergence of the quadruple mutant when maximal recombination between all loci is allowed (Fig. S6B). When only inter-epitope recombination occurs, recombination remains beneficial (with even a slightly increased magnitude) over most of the range, but the inhibitory effect seen at high f_e becomes stronger.

As there is uncertainty in the parameters of CTL dynamics, we also consider different values of the CTL proliferation rate, g (see

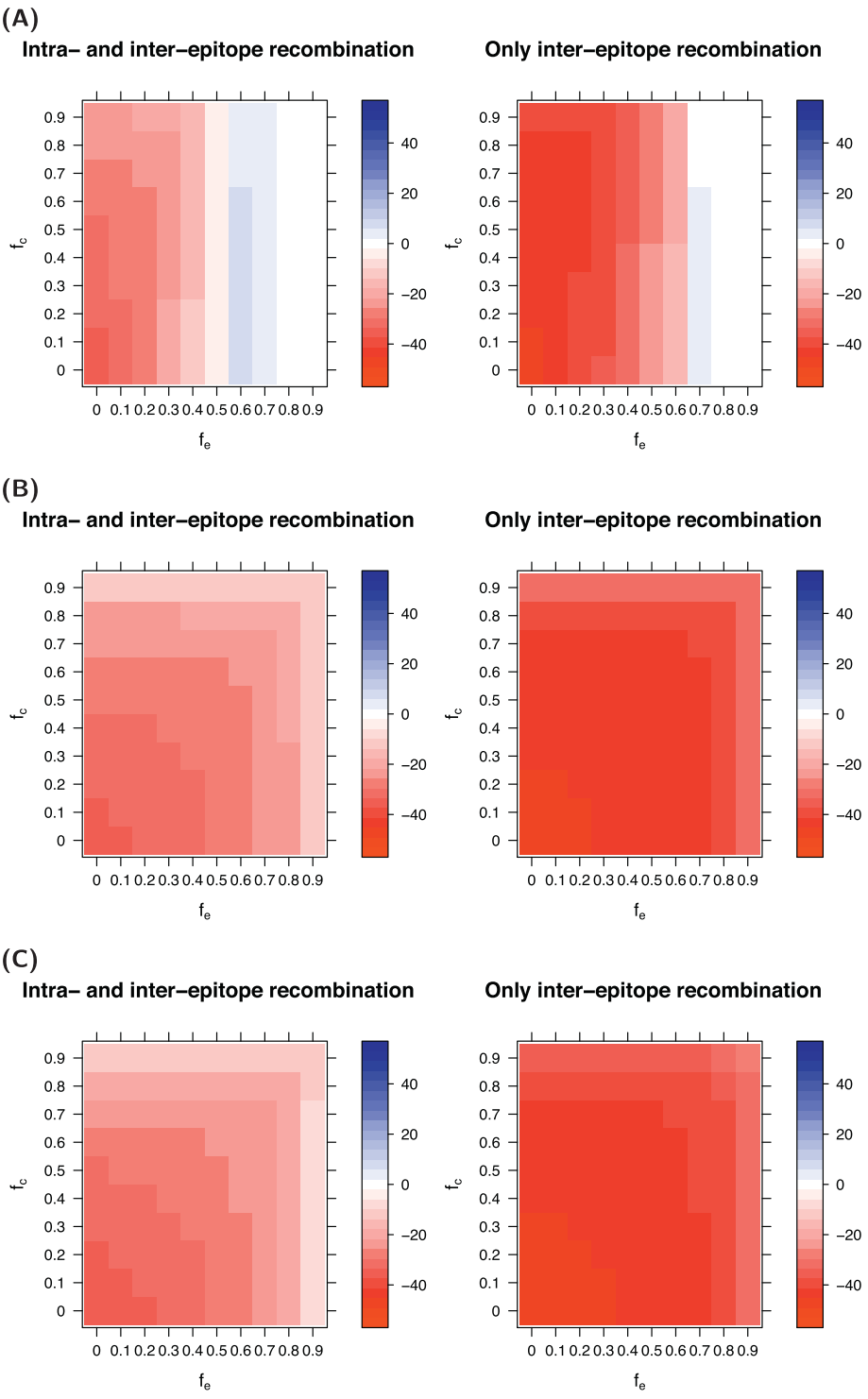

Fig. 4. Escape and reversion at two epitopes. We set the intrinsic fitness of the founder strain (f_{wt}) to 1 and the adapted strain (f_{ec}) to 0.9 throughout, and vary the intrinsic fitness associated with escape (f_e) and compensatory (f_c) mutations. The relative contribution of immune killing is $\epsilon = 0.9$. Other parameter values are listed in Table 1. The percentage change in the expected waiting time for the emergence of the fully adapted strain with versus without recombination ($R^*100\%$) when (A) both epitopes are subject to immune escape, (B) both subject to reversion, (C) one subject to escape and the other to reversion. The left panel illustrates the case when recombination can occur both within and between epitopes. The right panel illustrates the case when recombination can occur between epitopes but not within epitopes.

Text S1). Increasing g implies a higher CTL population before the start of escape. (Note that keeping ϵ constant requires a correspondingly lower per capita CTL killing rate (k_{CTL}).) The net CTL killing rate rises with the CTL population and eventually saturates. Following the emergence of the strain carrying escape and compensatory mutations at one epitope, it is targeted by a smaller population of CTLs that continue to recognize the other epitope. With higher g, this reduced population leads to a much smaller decline in the net CTL killing rate due to saturation effects, and hence slower growth, than with smaller g (Fig. S8; compare with Fig. 5). The effect of recombination when g is increased with fixed ϵ (Fig. S9A) is similar to that when ϵ is lowered with fixed g (Fig. S6B), which also leads to lower growth due to greater non-specific killing (Fig. S7; compare with Figs. 5 and S8). Correspondingly, decreasing

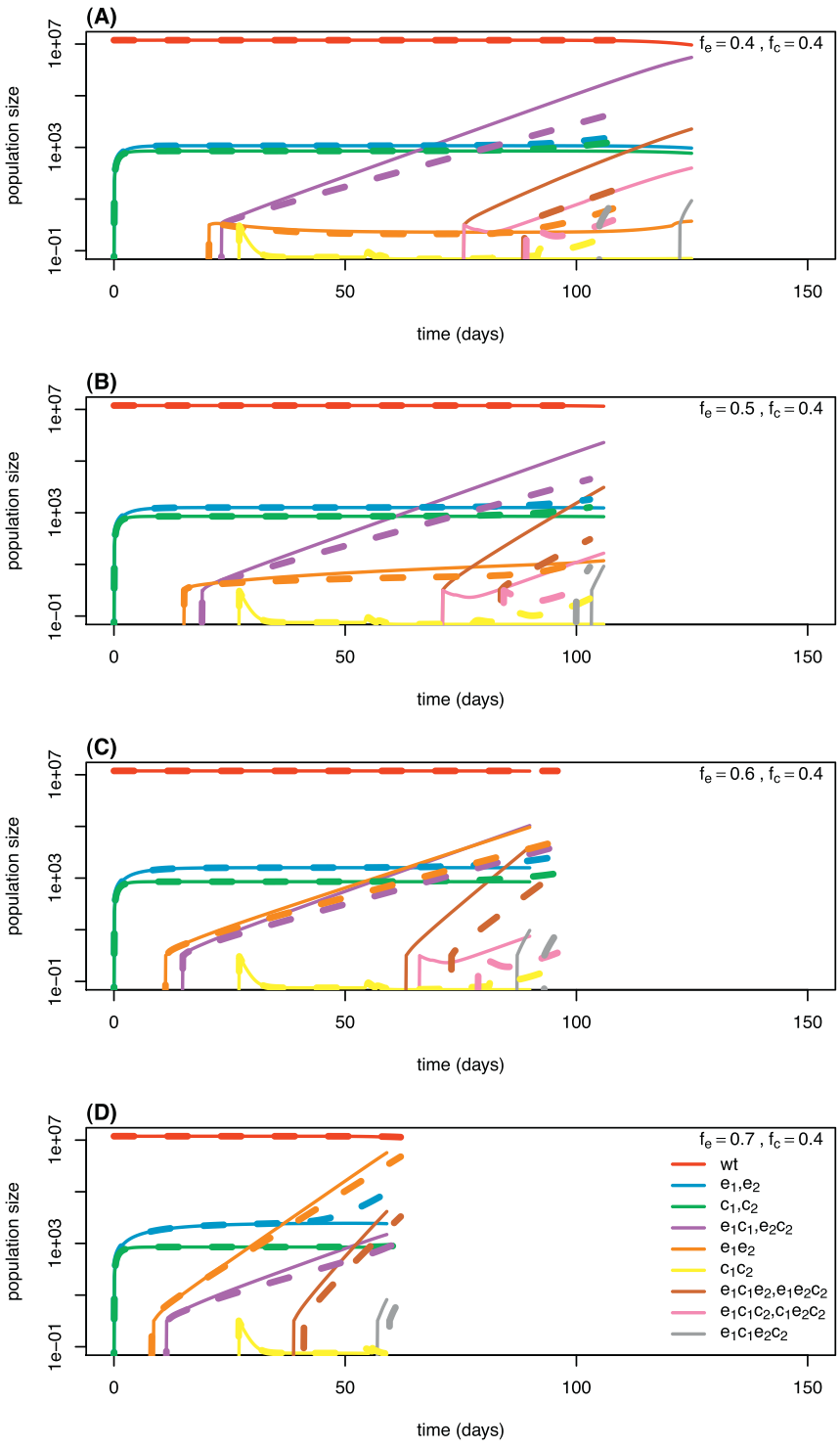

Fig. 5. Dynamics of escape at two epitopes. Time evolution of populations of various singly infected cells for selected intrinsic fitness values of the escape mutation, as indicated in each panel, with maximal recombination between all loci (dashed lines) or without recombination (solid lines). The other parameter values employed are the same as in Fig. 4A. Note that by symmetry the two epitopes are interchangeable, and thus the corresponding strains (e.g., e_1 and e_2, or e_1c_1 and e_2c_2) have identical dynamics.

g renders recombination beneficial over larger ranges of f_e and f_c (Figs. S9B and S9C).

3.2.2. *Reversion at both epitopes*

Reversion of *ec* at any given epitope involves passing through a fitness valley (i.e. a single mutation, here the loss of *e* or *c* alone, reduces fitness) to a higher peak on the other side (i.e. two mutations, here the loss of both *e* and *c*, increases fitness relative to the founder). Thus reversion at two epitopes involves adaptation on a fitness landscape similar to the case of escape at two epitopes with low f_e (creating a fitness valley at each epitope) and high ϵ (implying elevated fitness of the double mutant at each epitope). The outcome is thus also similar: recombination uniformly accelerates reversion at two epitopes (Fig. 4B). Starting with the quadruple mutant ($e_1c_1e_2c_2$), the triple mutants are produced first. The double mutants are produced next, of which the e_1c_1 and e_2c_2 strains grow

due to elevated fitness (Fig. S10). Recombination with the quadruple mutant slows their rise by breaking them into their precursor triple mutants. At the same time, recombination can bring the two double mutants together and produce the wild-type, completing the reversion process. The latter effect of recombination predominates, resulting in the net benefit of recombination to reversion at two symmetrical epitopes.

Again, if recombination only occurs between epitopes and is thus unable to break the escape and compensatory mutations within an epitope, its inhibitory effect is suppressed and a greater acceleration of reversion is observed. Further, as all strains are assumed equally susceptible to immune killing in the reversion process, altering ϵ or g (not shown) alters the influence of recombination only marginally (Fig. S11).

3.2.3. Escape at one and reversion at the other epitope

When escape at one and reversion at the other epitope are involved, the effects of recombination appear as a composite of the effects observed above with escape and reversion at both epitopes. Now, supposing epitope 1 is targeted and epitope 2 is reverting, adaptation begins with the founder strain e_2c_2, with both escape and compensatory mutations on the reverting epitope and none on the targeted epitope. We follow the adaptation until the emergence of the highly adapted e_1c_1 strain where both escape and reversion are complete. Although we assume that the intrinsic fitness landscape is identical for the two epitopes, the targeted epitope has superimposed on this landscape pressure from the immune system, and thus the overall fitness landscapes of the two epitopes differ. Consequently, evolution at the two epitopes does not proceed symmetrically, unlike the cases above. For instance, the single mutants e_1 and e_2 now have different fitnesses and thus follow different dynamics, introducing an additional dimension of inter-epitope competition.

The reverting epitope undergoes adaptation on a fitness landscape that contains a valley. Consequently, single mutations at this epitope remain at low frequencies, whereas the *wt* once it appears can outcompete the founder. The targeted epitope, on the other hand, either must pass through a fitness valley or may access a purely uphill trajectory depending on f_e. In the former case, the $e_1e_2c_2$ and $c_1e_2c_2$ strains arise but remain at low frequencies, whereas in the latter case, $e_1e_2c_2$ can grow. An additional mutation gives rise to the $e_1c_1e_2c_2$ strain. The subsequent growth of the *wt* and $e_1c_1e_2c_2$ are influenced by recombination and the fitness advantage gained by immune escape. The *wt* always enjoys a fitness advantage relative to the founder strain, whereas the quadruple mutant has a fitness advantage only when ϵ is sufficiently high. The growth of both strains can be inhibited by recombination with the founder strain, but recombination with one another can produce the highly adapted e_1c_1 strain (Fig. S12). Accordingly, when ϵ is large, recombination accelerates adaptation (Fig. 4C). As ϵ decreases, the benefit of recombination to adaptation diminishes (Fig. S13A), and eventually recombination becomes detrimental (Fig. S13B). When recombination within epitopes does not occur, the *ec* combinations are not split, so that recombination always benefits adaptation (Fig. S13A and S13B). In all cases, the influence of recombination is intermediate to that in the scenarios involving escape or reversion at both epitopes.

4. Discussion

Early adaptation of HIV-1 following its transmission typically combines the acquisition of escape mutations at epitopes targeted by CTLs in the new host and the reversion of mutations in the transmitted viral genome at epitopes that were targets in the donor but not in the new host. Despite the large mutation and turnover rates of HIV, mutation may prove inadequate as the sole driving force for this adaptation: the potential fitness loss upon acquiring escape or compensatory mutations alone, together with clonal interference among strains that have adapted at different epitopes, inhibits the accumulation of the requisite adaptive mutations at multiple loci. Recombination may then play an important role in accelerating this adaptation, under circumstances that we clarified here.

Using a detailed model of viral dynamics combining immune response by CTLs and viral evolution driven by mutation, recombination and selection, we investigated the expected waiting time for the appearance of the highly adapted strain, carrying all of the requisite mutations, in the presence and absence of recombination. We find, interestingly, that the dominant pathway of adaptation can differ substantially based on the underlying fitness landscape and that recombination can have opposite net influences on adaptation depending on the pathway involved. We summarize our findings using the scenario where adaptation involves immune escape at two identical epitopes (Fig. 6). If escape mutations carry a significant fitness penalty, then adaptation at each epitope passes through a fitness valley on the way to the *ec* double mutants. Subsequently, recombination can separate the double mutants into their constituent *e* and *c* single mutants, slowing adaptation. At the same time, it can bring the two double mutants together to yield the highly adapted quadruple mutant. So long as immune escape at each epitope confers a sufficient benefit, so that the *ec* double mutants reach high enough frequency for substantial coinfection together, the latter effect dominates and recombination accelerates adaptation (Fig. 6A). If however immune escape at one epitope does not confer a sufficient benefit (which is the case if CTL killing contributes little to total infected cell death), the *ec* double mutants remain at lower frequency and coinfection is primarily with the wild type, so that the former effect dominates and recombination decelerates adaptation. Of course, if the *e* and *c* mutations on each epitope are proximally located and hence tightly linked, then recombination is ineffective in separating the double mutants, so that its overall effect tends to be beneficial (or less detrimental).

If escape mutations have an overall fitness advantage alone (even without compensation), adaptation at each epitope can proceed on a purely uphill trajectory. The *e* mutants then grow, yielding the double mutant e_1e_2. Adaptation now proceeds on the background of the e_1e_2 strain rather than the *ec* strains as above. Since c_1c_2 is of low fitness and thus rare, recombination bringing together e_1e_2 and c_1c_2 is negligible. The predominant effect of recombination is thus to separate the e_1e_2 double mutant, delaying the emergence of triple mutants and hence further adaptation (Fig. 6B). This delay outweighs the subsequent possibility for recombination to accelerate adaptation by bringing the triple mutants together.

To summarize, when adaptation passes through a fitness valley to a sufficiently high fitness gain at each epitope, recombination acting between epitopes brings well-adapted pairs of mutations together and has the net effect of accelerating adaptation. When an uphill trajectory is accessible at each epitope, the primary role of recombination is to separate mutations on the dominant intermediate double mutant, delaying subsequent adaptation. This difference in the dominant adaptation pathway underlies the different roles of recombination we observed for immune escape versus reversion at multiple symmetrical epitopes. While reversion occurs on the former type of landscape, escape could fall into either scenario, depending on the intrinsic fitness cost of escape mutations (which is observed to vary widely; see e.g. Troyer et al., 2009) and the extent to which escape reduces total infected cell death (which is not well known). Our results can also be interpreted in terms of the distances between loci that maximize rate of

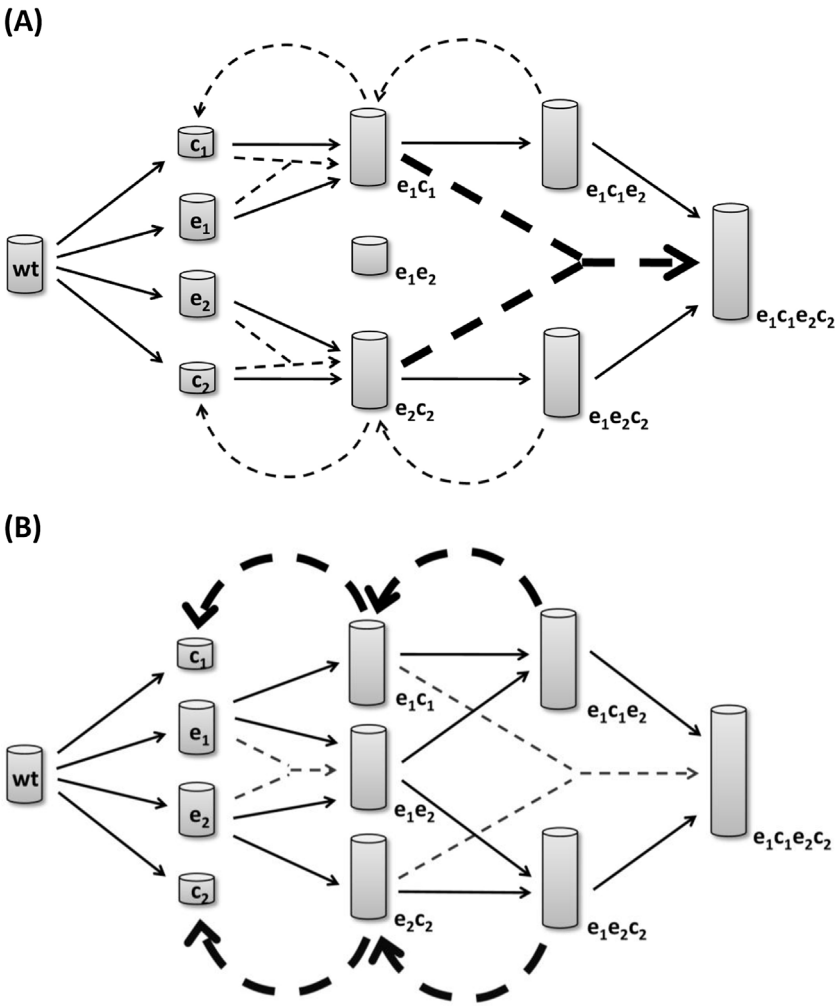

Fig. 6. Illustration of the influence of recombination on early HIV-1 adaptation, specifically the scenario of escape at two epitopes. The height of each bar indicates the overall fitness (F) of the corresponding strain. (A) Case where F_e is small but F_{ec} is high (due to low f_e while ϵ is sufficiently high) and a fitness valley at each epitope must be crossed (cf. dynamics in Fig. 5A). (B) Case where F_e is large (due to high f_e) and adaptation can proceed by a purely uphill fitness trajectory at each epitope (cf. dynamics in Fig. 5C). Only major pathways of mutation (solid arrows) and recombination (dashed arrows) are shown, with the most significant recombination pathways highlighted with thicker arrows. For simplicity, recombination that breaks up combinations of mutations is shown with an arrow only from the higher- to lower-order mutant, i.e. not illustrating the other 'parent' genome involved in the recombination event. Less fit double mutants (e_1c_2, c_1e_2, c_1c_2) and triple mutants ($e_1c_1c_2, c_1e_2c_2$) which play a minimal role in the dynamics are not shown.

adaptation: while linkage between various escape mutations may be beneficial or detrimental, depending as described on the fitness landscape, tight linkage between a given escape mutation and its associated compensatory mutation (which exhibits positive epistasis) appears to be generally beneficial for multi-epitope adaptation. It appears that associated escape and compensatory mutations are often close together, though not necessarily within the same epitope (Brockman et al., 2007; Crawford et al., 2007; Schneidewind et al., 2007; Wright et al., 2012; Liu et al., 2014). Generally speaking, our study predicts that both fitness effects (costs and benefits) and linkage patterns among loci involved in adaptation are crucial parameters. Further empirical work to quantify these parameters could clarify which scenario is most frequently faced by HIV, and thus help resolve the ambiguous evolutionary advantage of undergoing recombination.

In the present study, we have restricted our analysis to a maximum of two epitopes (thus four loci). With two epitopes, the effects of recombination in both bringing together and breaking up mutations, thus affecting both appearance and growth of intermediate strains, already become evident, while the system still remains amenable to efficient computation and a thorough understanding of dynamics. Our model is nonetheless constructed more generally and can be applied to adaptation at more than two epitopes. We expect computational speed, however, to become prohibitive beyond three or four epitopes, in which case it would be worthwhile to investigate approximations to certain processes in the model. If more reverting epitopes with similar fitness effects are considered, we anticipate, based on our one- and two-epitope results, that recombination is likely to expedite adaptation. If more targeted (escaping) epitopes are considered, the influence of recombination would depend on the fitness landscape, as determined by both the intrinsic replicative capacity and the relative contribution of immune killing. Unequal fitness effects across adapting epitopes could add further nuances to these results. A previous study, using a less detailed model with a heuristic description of recombination, considered up to 15 targeted epitopes and predicted that recombination has an increasingly accelerating effect on adaptation as the number of epitopes increases (Mostowy et al., 2011). However, averaging results over randomly drawn fitness costs of escape in that study precluded such a detailed understanding of the role of the fitness landscape as we achieved here. We recognize though that as the number of epitopes increases, it may become increasingly important to account for fitness interactions among epitopes. More realistic fitness landscapes of HIV-1,

including epistatic interactions, are beginning to be unraveled (Hinkley et al., 2011; Ferguson et al., 2013).

Some modelling studies of SIV infection in rhesus macaques allude to antibody dynamics during acute infection (Vaidya et al., 2010). While antibodies appear within the first few weeks of infection, they do not seem to have a significant effect on viremia. Antibodies with significant neutralizing activity emerge only after several months to years past the acute infection stage (Overbaugh and Morris, 2012). In contrast, a median of eight epitopes are targeted by CTLs within three months from the onset of symptoms (Turnbull et al., 2009), and CTL responses are strongly associated with several orders of magnitude decline in viremia (Walker and McMichael, 2012). We have therefore focused on adaptation in the context of CTL responses. Furthermore, we investigated a relatively simplistic immune response, in which we assumed that all effector clones behave identically and begin at their steady-state levels. We did not deal with different CTL avidities, leading to immunodominance patterns, or different killing efficiencies at different epitopes. Previous modelling studies have indicated that asymmetries among effector cell clones can affect escape dynamics (Althaus and De Boer, 2008; Mostowy et al., 2011; van Deutekom et al., 2013), and that the more realistic scenario in which escape mutations may only partially abrogate CTL killing affects the rise of these mutations (Batorsky et al., 2014). Furthermore, in the two-epitope studies, we assumed the intrinsic fitness effects at the two epitopes to be identical. This assumption helps in delineating the influence of recombination on multi-locus adaptation while keeping the complexities of clonal interference to a minimum. Fitness effects, however, are unlikely to be identical at any two epitopes. This asymmetry is considered in our study in the case where one epitope is escaping and the other is reverting, such that the two epitopes then have different overall fitness landscapes. Further exploration of the effects of both epitope asymmetries and more complex immune responses using our model in its general form could provide interesting directions for future work. Finally, CTL killing efficiencies can be compromised by T cell exhaustion, a phenomenon of great current interest (Wherry and Kurachi, 2015). Following the formalism proposed by Johnson et al. (2011), where the level of exhaustion is determined by the level of cumulative antigenic stimulation, we found over a range of parameter values encompassing those employed in the latter study that exhaustion did not alter the influence of recombination we predict (not shown). Thus, it appears that the general effects of recombination we found are fairly robust. However, a full exploration of the effects of different models of CTL response are beyond the scope of the present study.

Our model in its present form is fully deterministic, allowing ready exploration of parameter space and rendering it amenable to data analysis. Unlike many deterministic models, however, it has the advantage of capturing delays in the emergence of mutant strains in a finite population, which enables it to capture the Fisher–Muller effect. Indeed, a similar formalism in the context of drug resistance, based on which the current study is built, quantitatively captured the in vitro development of resistance to tipranavir, an HIV protease inhibitor with a large genetic barrier (Arora and Dixit, 2009). A limitation of the formalism is that stochastic variations around the mean are ignored, though these may contribute substantially to immune escape and control (Read et al., 2012; Kessinger et al., 2013). In particular, in a stochastic setting, strains requiring the same number of mutational steps may appear at different times, and strains with the same expected fitness may vary in their growth. This will affect the availability of relevant strains for recombination, even if we continue to assume identical fitness effects across epitopes.

While still removed from making specific clinical suggestions, our results have general implications for understanding and managing HIV. Firstly, the substantial effect of recombination on adaptation rate that we predict indicates that linkage patterns among adapting loci should be taken into account in the design of a T-cell vaccine targeting multiple epitopes. Though the focus on targeting conserved sites where mutations carry large costs (Martinez-Picado et al., 2006; Schneidewind et al., 2007; Goepfert et al., 2008; Dahirel et al., 2011) is justified, an additional moderating factor could be the distances between predicted escape mutations as well as their associated compensatory mutations. Secondly, the influence of recombination has implications for quantifying the timing and rate of escape from patient data, currently an active research area (Asquith et al., 2006; Ganusov et al., 2006, 2011, 2013; Kessinger et al., 2013; Pandit and de Boer, 2014; Garcia and Regoes, 2014). Most models used for this purpose have treated each escape locus independently (Asquith et al., 2006; Ganusov et al., 2006, 2011). Recent results indicate that clonal interference is expected to bias estimates of escape rates based on individual sites (Ganusov et al., 2013; Kessinger et al., 2013; Pandit and de Boer, 2014; Garcia and Regoes, 2014), and an inference method for multi-locus data assuming complete linkage (no recombination) has recently been developed (Kessinger et al., 2013). Our results clarify the potentially significant effect of recombination on the timing at which new strains arise and how fast they spread. With growing availability of haplotype data that preserves information about linkage of escape mutations (e.g. Henn et al., 2012), our work emphasizes the need for further development of inference methods that take recombination into account.

Acknowledgements

We acknowledge the Indo-Swiss Joint Research Programme (Personnel Exchange Programme), funded by the Swiss National Science Foundation and the Department of Science and Technology, Government of India, for supporting this collaboration between our groups at the Indian Institute of Science and the ETH Zürich. Support is also acknowledged from the Wellcome Trust/DBT India Alliance Senior Fellowship (IA/S/14/1/501307). We thank Christian Althaus, Roland Regoes, and Victor Garcia for helpful discussions during the development and parameterization of our model, Shibin Mathew for help with the formulation of the recombination probabilities, Gabriel Leventhal for assistance in debugging our code, Pranesh Padmanabhan for help with the illustration in Fig. 6, and Roland Regoes for comments on a draft of this manuscript. We also thank two anonymous reviewers, whose insightful comments spurred improvements to this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.epidem.2015.09.001

References

Althaus, C.L., Bonhoeffer, S., 2005. Stochastic interplay between mutation and recombination during the acquisition of drug resistance mutations in human immunodeficiency virus type 1. J. Virol. 79, 13572–13578.

Althaus, C.L., De Boer, R.J., 2008. Dynamics of immune escape during HIV/SIV infection. PLoS Comput. Biol. 4, e1000103.

Arora, P., Dixit, N.M., 2009. Timing the emergence of resistance to anti-HIV drugs with large genetic barriers. PLoS Comput. Biol. 5, e1000305.

Asquith, B., Edwards, C.T., Lipsitch, M., McLean, A.R., 2006. Inefficient cytotoxic T lymphocyte-mediated killing of HIV-1-infected cells in vivo. PLoS Biol. 4, e90.

Batorsky, R., Kearney, M.F., Palmer, S.E., Maldarelli, F., Rouzine, I.M., et al., 2011. Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. Proc. Natl. Acad. Sci. U.S.A. 108, 5661–5666.

Batorsky, R., Sergeev, R.A., Rouzine, I.M., 2014. The route of HIV escape from immune response targeting multiple sites is determined by the cost-benefit tradeoff of escape mutations. PLoS Comput. Biol. 10, e1003878.

Beauchemin, C., Dixit, N.M., Perelson, A.S., 2007. Characterizing T cell movement within lymph nodes in the absence of antigen. J. Immunol. 178, 5505–5512.
Bocharov, G., Ford, N.J., Edwards, J., Breinig, T., Wain-Hobson, S., et al., 2005. A genetic-algorithm approach to simulating human immunodeficiency virus evolution reveals the strong impact of multiply infected cells and recombination. J. Gen. Virol. 86, 3109–3118.
Boerlijst, M., Bonhoeffer, S., Nowak, M., 1996. Viral quasi-species and recombination. Proc. R. Soc. Lond. B 263, 1577–1584.
Bonhoeffer, S., Chappey, C., Parkin, N.T., Whitcomb, J.M., Petropoulos, C.J., 2004. Evidence for positive epistasis in HIV-1. Science 306, 1547–1550.
Borghans, J.A., de Boer, R.J., Segel, L.A., 1996. Extending the quasi-steady state approximation by changing variables. Bull. Math. Biol. 58, 43–63.
Bretscher, M.T., Althaus, C.L., Muller, V., Bonhoeffer, S., 2004. Recombination in HIV and the evolution of drug resistance: for better or for worse? Bioessays 26, 180–188.
Brockman, M.A., Schneidewind, A., Lahaie, M., Schmidt, A., Miura, T., et al., 2007. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J. Virol. 81, 12608–12618.
Budhu, S., Loike, J.D., Pandolfi, A., Han, S., Catalano, G., et al., 2010. CD8+ T cell concentration determines their efficiency in killing cognate antigen-expressing syngeneic mammalian cells in vitro and in mouse tissues. J. Exp. Med. 207, 223–235.
Carvajal-Rodriguez, A., Crandall, K.A., Posada, D., 2007. Recombination favors the evolution of drug resistance in HIV-1 during antiretroviral therapy. Infect. Genet. Evol. 7, 476–483.
Chen, B., Gandhi, R., Baltimore, D., 1996. CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef. J. Virol. 70, 6044–6053.
Chen, J., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C.E., et al., 2009. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. Proc. Natl. Acad. Sci. U.S.A. 106, 13535–13540.
Chopera, D., Woodman, Z., Mlisana, K., Mlotshwa, M., Martin, D., et al., 2008. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog. 4, e1000033.
Coffin, J.M., 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267, 483–489.
Crawford, H., Prado, J.G., Leslie, A., Hue, S., Honeyborne, I., et al., 2007. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. J. Virol. 81, 8346–8351.
Crawford, H., Lumm, W., Leslie, A., Schaefer, M., Boeras, D., et al., 2009. Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. J. Exp. Med. 206, 909–921.
Crawford, H., Matthews, P.C., Schaefer, M., Carlson, J.M., Leslie, A., et al., 2011. The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms. J. Virol. 85, 1384–1390.
da Silva, J., 2012. The dynamics of HIV-1 adaptation in early infection. Genetics 190, 1087–1099.
Dahirel, V., Shekhar, K., Pereyra, F., Miura, T., Artyomov, M., et al., 2011. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc. Natl. Acad. Sci. U.S.A. 108, 11530–11535.
De Boer, R.J., 2007. Understanding the failure of CD8+ T-cell vaccination against simian/human immunodeficiency virus. J. Virol. 81, 2838–2848.
Dixit, N.M., 2008. Modeling HIV infection dynamics: the role of recombination in the development of drug resistance. Future HIV Ther. 2, 375–388.
Ferguson, A.L., Mann, J.K., Omarjee, S., Ndung'u, T., Walker, B.D., et al., 2013. Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. Immunity 38, 606–617.
Fernandez, C.S., Stratov, I., De Rose, R., Walsh, K., Dale, C.J., et al., 2005. Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost. J. Virol. 79, 5721–5731.
Fischer, W., Ganusov, V.V., Giorgi, E.E., Hraber, P.T., Keele, B.F., et al., 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS ONE 5, e12303.
Fraser, C., 2005. HIV recombination: what is the impact on antiretroviral therapy? J. R. Soc. Interface 2, 489–503.
Friedrich, T.C., Dodds, E.J., Yant, L.J., Vojnov, L., Rudersdorf, R., et al., 2004. Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nat. Med. 10, 275–281.
Gadhamsetty, S., Dixit, N.M., 2010. Estimating frequencies of minority nevirapine-resistant strains in chronically HIV-1-infected individuals naive to nevirapine by using stochastic simulations and a mathematical model. J. Virol. 84, 10230–10240.
Gadhamsetty, S., Maree, A.F., Beltman, J.B., de Boer, R.J., 2014. A general functional response of cytotoxic T lymphocyte-mediated killing of target cells. Biophys. J. 106, 1780–1791.
Ganusov, V.V., De Boer, R.J., 2006. Estimating costs benefits of CTL escape mutations in SIV/HIV infection. PLoS Comput. Biol. 2, e24.
Ganusov, V.V., Goonetilleke, N., Liu, M.K., Ferrari, G., Shaw, G.M., et al., 2011. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. J. Virol. 85, 10518–10528.
Ganusov, V.V., Barber, D.L., De Boer, R.J., 2011. Killing of targets by CD8 T cells in the mouse spleen follows the law of mass action. PLoS ONE 6 (1), e15959.
Ganusov, V.V., Neher, R.A., Perelson, A.S., 2013. Mathematical modeling of escape of HIV from cytotoxic T lymphocyte responses. J. Stat. Mech., P01010.
Gao, F., Wang, D., 2007. Minor-drug-resistant HIV populations and treatment failure. Future Virol. 2, 293–302.
Garcia, V., Regoes, R.R., 2014. The effect of interference on the CD8(+) T cell escape rates in HIV. Front. Immunol. 5, 661.
Gerrish, P.J., Lenski, R.E., 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102-103, 127–144.
Gheorghiu-Svirschevski, S., Rouzine, I.M., Coffin, J.M., 2007. Increasing sequence correlation limits the efficiency of recombination in a multisite evolution model. Mol. Biol. Evol. 24, 574–586.
Goepfert, P.A., Lumm, W., Farmer, P., Matthews, P., Prendergast, A., et al., 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J. Exp. Med. 205, 1009–1017.
Goonetilleke, N., Liu, M.K., Salazar-Gonzalez, J.F., Ferrari, G., Giorgi, E., et al., 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J. Exp. Med. 206, 1253–1272.
Graw, F., Regoes, R.R., 2009. Investigating CTL mediated killing with a 3D cellular automaton. PLoS Comput. Biol. 5, e1000466.
Hartl, D., Clark, A., 2007. Principles of Population Genetics. Sinauer Associates, Sunderland, MA.
Henn, M.R., Boutwell, C.L., Charlebois, P., Lennon, N.J., Power, K.A., et al., 2012. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog. 8, e1002529.
Hill, W.G., Robertson, A., 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8, 269–294.
Hinkley, T., Martins, J., Chappey, C., Haddad, M., Stawiski, E., et al., 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat. Genet. 43, 487–489.
Johnson, P.L., Kochin, B.F., McAfee, M.S., Stromnes, I.M., Regoes, R.R., et al., 2011. Vaccination alters the balance between protective immunity, exhaustion, escape, and death in chronic infections. J. Virol. 85, 5565–5570.
Josefsson, L., King, M.S., Makitalo, B., Brannstrom, J., Shao, W., et al., 2011. Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. Proc. Natl. Acad. Sci. U.S.A. 108, 11199–11204.
Josefsson, L., Palmer, S., Faria, N.R., Lemey, P., Casazza, J., et al., 2013. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. PLoS Pathog. 9, e1003432.
Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., et al., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U.S.A. 105, 7552–7557.
Kelleher, A.D., Long, C., Holmes, E.C., Allen, R.L., Wilson, J., et al., 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. 193, 375–386.
Kessinger, T.A., Perelson, A.S., Neher, R.A., 2013. Inferring HIV escape rates from multi-locus genotype data. Front. Immunol. 4, 252.
Kim, Y., Orr, H.A., 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher–Muller model. Genetics 171, 1377–1386.
Kouyos, R.D., Otto, S.P., Bonhoeffer, S., 2006. Effect of varying epistasis on the evolution of recombination. Genetics 173, 589–597.
Kouyos, R.D., Silander, O.K., Bonhoeffer, S., 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol. Evol. (Amst.) 22, 308–315.
Kouyos, R.D., Fouchet, D., Bonhoeffer, S., 2009. Recombination and drug resistance in HIV: population dynamics and stochasticity. Epidemics 1, 58–69.
Leslie, A.J., Pfafferott, K.J., Chetty, P., Draenert, R., Addo, M.M., et al., 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat. Med. 10, 282–289.
Leviyang, S., 2013. Computational inference methods for selective sweeps arising in acute HIV infection. Genetics 194, 737–752.
Levy, D.N., Aldrovandi, G.M., Kutsch, O., Shaw, G.M., 2004. Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. U.S.A. 101, 4204–4209.
Liu, M.K., Hawkins, N., Ritchie, A.J., Ganusov, V.V., Whale, V., et al., 2013. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. J. Clin. Invest. 123 (1), 380–393.
Liu, D., Zuo, T., Hora, B., Song, H., Kong, W., et al., 2014. Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation. Retrovirology 11, 101.
Martinez-Picado, J., Prado, J.G., Fry, E.E., Pfafferott, K., Leslie, A., et al., 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J. Virol. 80, 3617–3623.
McMichael, A.J., Borrow, P., Tomaras, G.D., Goonetilleke, N., Haynes, B.F., 2010. The immune response during acute HIV-1 infection: clues for vaccine development. Nat. Rev. Immunol. 10, 11–23.
Miller, M.J., Safrina, O., Parker, I., Cahalan, M.D., 2004. Imaging the single cell dynamics of CD4+ T cell activation by dendritic cells in lymph nodes. J. Exp. Med. 200, 847–856.
Miura, T., Brockman, M.A., Brumme, Z.L., Brumme, C.J., Pereyra, F., et al., 2009a. HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J. Virol. 83, 140–149.
Miura, T., Brockman, M.A., Schneidewind, A., Lobritz, M., Pereyra, F., et al., 2009b. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select

for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte recognition. J. Virol. 83, 2743–2755.
Miura, T., Brumme, Z.L., Brockman, M.A., Rosato, P., Sela, J., et al., 2010. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. J. Virol. 84, 7581–7591.
Moradigaravand, D., Kouyos, R., Hinkley, T., Haddad, M., Petropoulos, C.J., et al., 2014. Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. PLoS Genet. 10 (6), e1004439.
Mostowy, R., Kouyos, R.D., Fouchet, D., Bonhoeffer, S., 2011. The role of recombination for the coevolutionary dynamics of HIV and the immune response. PLoS ONE 6, e16052.
Moutouh, L., Corbeil, J., Richman, D.D., 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. Proc. Natl. Acad. Sci. U.S.A. 93, 6106–6111.
Neher, R.A., Leitner, T., 2010. Recombination rate and selection strength in HIV intra-patient evolution. PLoS Comput. Biol. 6, e1000660.
Nowak, M., May, R.M., 2000. Virus Dynamics, Mathematical Principles of Immunology and Virology. Oxford University Press, New York.
Otto, S.P., Lenormand, T., 2002. Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3, 252–261.
Overbaugh, J., Morris, L., 2012. The antibody response against HIV-1. Cold Spring Harb. Perspect. Med. 2, a007039.
Pandit, A., de Boer, R.J., 2014. Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. Retrovirology 11, 56.
Pennings, P.S., 2012. Standing genetic variation and the evolution of drug resistance in HIV. PLoS Comput. Biol. 8, e1002527.
Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., Ho, D.D., 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271, 1582–1586.
Piguet, V., Gu, F., Foti, M., Demaurex, N., Gruenberg, J., et al., 1999. Nef-induced CD4 degradation: a diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of beta-COP in endosomes. Cell 97, 63–73.
Read, E.L., Tovo-Dwyer, A.A., Chakraborty, A.K., 2012. Stochastic effects are important in intrahost HIV evolution even when viral loads are high. Proc. Natl. Acad. Sci. U.S.A. 109, 19727–19732.
Rhodes, T., Wargo, H., Hu, W.S., 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. J. Virol. 77, 11193–11200.
Rouzine, I.M., Coffin, J.M., 2005. Evolution of human immunodeficiency virus under selection and weak recombination. Genetics 170, 7–18.
Rouzine, I.M., Coffin, J.M., 2010. Multi-site adaptation in the presence of infrequent recombination. Theor. Popul. Biol. 77, 189–204.
Salazar-Gonzalez, J.F., Salazar, M.G., Keele, B.F., Learn, G.H., Giorgi, E.E., et al., 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J. Exp. Med. 206, 1273–1289.
Schneidewind, A., Brockman, M.A., Yang, R., Adam, R.I., Li, B., et al., 2007. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J. Virol. 81, 12382–12393.
Schneidewind, A., Brockman, M.A., Sidney, J., Wang, Y.E., Chen, H., et al., 2008. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. J. Virol. 82, 5594–5605.
Schneidewind, A., Brumme, Z.L., Brumme, C.J., Power, K.A., Reyor, L.L., et al., 2009. Transmission and long-term stability of compensated CD8 escape mutations. J. Virol. 83, 3993–3997.
Suryavanshi, G.W., Dixit, N.M., 2007. Emergence of recombinant forms of HIV: dynamics and scaling. PLoS Comput. Biol. 3, 2003–2018.
Troyer, R.M., McNevin, J., Liu, Y., Zhang, S.C., Krizan, R.W., et al., 2009. Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. PLoS Pathog. 5, e1000365.
Turnbull, E.L., Wong, M., Wang, S., Wei, X., Jones, N.A., et al., 2009. Kinetics of expansion of epitope-specific T cell responses during primary HIV-1 infection. J. Immunol. 182, 7131–7145.
Vaidya, N.K., Ribeiro, R.M., Miller, C.J., Perelson, A.S., 2010. Viral dynamics during primary simian immunodeficiency virus infection: effect of time-dependent virus infectivity. J. Virol. 84, 4302–4310.
van Deutekom, H.W., Wijnker, G., de Boer, R.J., 2013. The rate of immune escape vanishes when multiple immune responses control an HIV infection. J. Immunol. 191, 3277–3286.
Vijay, N.N.V., Vasantika, Ajmani, R., Perelson, A.S., Dixit, N.M., 2008. Recombination increases human immunodeficiency virus fitness, but not necessarily diversity. J. Gen. Virol. 89, 1467–1477.
Walker, B., McMichael, A., 2012. The T-cell response to HIV. Cold Spring Harb. Perspect. Med., 2.
Wherry, E.J., Kurachi, M., 2015. Molecular and cellular insights into T cell exhaustion. Nat. Rev. Immunol. 15, 486–499.
Wodarz, D., 2014. Modeling T cell responses to antigenic challenge. J. Pharmacokinet. Pharmacodyn. 41, 415–429.
Wright, J.K., Naidoo, V.L., Brumme, Z.L., Prince, J.L., Claiborne, D.T., et al., 2012. Impact of HLA-B*81-associated mutations in HIV-1 Gag on viral replication capacity. J. Virol. 86, 3193–3199.